

Surface and underwater target classification under limited sample sizes based on sound field elevation structure

Yixin Miao(苗艺馨), Jin Fu(付进), and Xue Wang(王雪)

Citation: Chin. Phys. B, 2025, 34 (**11**): 114301. DOI: 10.1088/1674-1056/ade389
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

A WKB method based on parabolic cylinder function for very-low-frequency sound propagation in deep ocean

Jian-Kang Zhan(詹建康), Sheng-Chun Piao(朴胜春), Li-Jia Gong(龚李佳), Yang Dong(董阳), Yong-Chao Guo(郭永超), and Guang-Xue Zheng(郑广学)
Chin. Phys. B, 2025, 34 (**3**): 034301. DOI: 10.1088/1674-1056/ada886

Acoustic propagation uncertainty in internal wave environments using an ocean-acoustic joint model

Fei Gao(高飞), Fanghua Xu(徐芳华), Zhenglin Li(李整林), Jixing Qin(秦继兴), and Qinya Zhang(章沁雅)
Chin. Phys. B, 2023, 32 (**3**): 034302. DOI: 10.1088/1674-1056/ac89dc

Wave mode computing method using the step-split Padé parabolic equation

Chuan-Xiu Xu(徐传秀) and Guang-Ying Zheng(郑广赢)
Chin. Phys. B, 2022, 31 (**9**): 094301. DOI: 10.1088/1674-1056/ac6940

Sound propagation in inhomogeneous waveguides with sound-speed profiles using the multimodal admittance method

Qi Li(李琪), Juan Liu(刘娟), Wei Guo(郭威)
Chin. Phys. B, 2020, 29 (**1**): 014303. DOI: 10.1088/1674-1056/ab5940

Acoustic radiation from the submerged circular cylindrical shell treated with active constrained layer damping

Yuan Li-Yun (袁丽芸), Xiang Yu (向宇), Lu Jing (陆静), Jiang Hong-Hua (蒋红华)
Chin. Phys. B, 2015, 24 (**12**): 124301. DOI: 10.1088/1674-1056/24/12/124301

Surface and underwater target classification under limited sample sizes based on sound field elevation structure

Yixin Miao(苗艺馨)[1,2,3], Jin Fu(付进)[4,†], and Xue Wang(王雪)[5]

[1] *National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin* 150001, *China*
[2] *Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin* 150001, *China*
[3] *College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin* 150001, *China*
[4] *School of Electronic Information Engineering, Beihang University, Beijing* 100094, *China*
[5] *Hangzhou Applied Acoustics Research Institute, Hangzhou* 310023, *China*

(Received 21 April 2025; revised manuscript received 8 June 2025; accepted manuscript online 11 June 2025)

Surface/underwater target classification is a key topic in marine information research. However, the complex underwater environment, coupled with the diversity of target types and their variable characteristics, presents significant challenges for classifier design. For shallow-water waveguides with a negative thermocline, a residual neural network (ResNet) model based on the sound field elevation structure is constructed. This model demonstrates robust classification performance even when facing low signal-to-noise ratios and environmental mismatches. Meanwhile, to address the reduced generalization ability caused by limited labeled acoustic data, an improved ResNet model based on unsupervised domain adaptation ("proposed UDA-ResNet") is further constructed. This model incorporates data on simulated elevation structures of the sound field to augment the training process. Adversarial training is employed to extract domain-invariant features from simulated and trial data. These strategies help reduce the negative impact caused by domain differences. Experimental results demonstrate that the proposed method shows strong surface/underwater target classification ability under limited sample sizes, thus confirming its feasibility and effectiveness.

Keywords: sound field elevation structure, surface/underwater target classification, limited sample size, unsupervised domain adaptation

PACS: 43.30.–k, 43.30.Bp, 43.60.Np, 07.05.Mh

DOI: 10.1088/1674-1056/ade389 **CSTR:** 32038.14.CPB.ade389

1. Introduction

In marine information research, the core task of surface/underwater target classification is to distinguish between surface and underwater targets by analyzing sonar-received underwater acoustic signals.[1] Current classification features mainly include target radiated noise characteristics,[2] target depth estimation,[3] and sound field features.[4] Among them, the sound field elevation structure is a sound field feature that varies with target depth. Due to its requirement for a smaller vertical array aperture, the sound field elevation structure is widely used as a classification feature.[5,6]

However, due to the complex underwater environment, the diversity of targets, and the variability of target characteristics, designing surface/underwater target classifiers remains a challenging task.[7,8] Existing classification methods often adopt the concept of binary detectors to construct binary classifiers.[9–11] These methods determine the class of a sample by comparing the estimated value of a classification feature with a predefined threshold. However, both the construction of classification features and the selection of classification thresholds largely depend on prior environmental information.[10] Mismatches in such information may lead to degraded classification performance. Other commonly used classifiers include the Bayes classifier and the linear classifier.[12] However, the Bayes classifier relies on accurate class conditional probability density, which is often difficult to obtain in practical applications. Linear classifiers assume that the decision boundary between classes is linear. However, in underwater acoustic target classification, class distributions are often nonlinear and complex, which makes linear classifiers less effective for such tasks.

As deep learning techniques have developed, applying their strong feature learning capabilities to surface/underwater target classification has gradually become a research focus. Choi *et al.* employed feedforward and convolutional neural networks to process cross-spectral density matrices, from which depth-related features were extracted to support surface and underwater target classification.[13] Luo *et al.* used deep convolutional neural networks to extract features from acoustic data and demonstrated the potential of deep learning in enhancing classification performance.[14] Zhou *et al.* proposed a compound convolutional neural network based on shared sparse features and deep belief networks to improve the classification of interference striation images.[15] Niu *et*

†Corresponding author. E-mail: fujin@buaa.edu.cn

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

al. used a deep residual network to predict source locations, which showed significantly better performance compared to matched field processing methods.[16] Additionally, some studies have introduced deep generative models for modeling three-dimensional sound speed fields, thereby assisting in downstream tasks such as surface and underwater target classification.[25,26] Overall, deep learning shows strong potential for increasing classification accuracy.

Deep learning methods generally require large amounts of training data. However, collecting sound field elevation structure samples is costly, and labeled data are often scarce.[18,19] These constraints make it challenging to train deep learning models under limited sample sizes. To address this issue, a UDA-ResNet model is constructed based on an unsupervised domain adaptation strategy,[20] in which simulated sound field elevation structure data are introduced to augment the training set. Through adversarial training, domain-invariant features common to the simulated and trial data are extracted, making the features obtained by the feature extractor indistinguishable in the domain discriminator. This approach improves the classification performance of the model under limited sample sizes.

2. Target depth dependence of sound field elevation structure samples

Consider a vertical array composed of N hydrophones, as shown in the figure below. The spacing between array elements is Δz. Taking the 0-th element located at depth z_0 as the reference, the depth of the n-th element is $z_0+n\Delta z$, where $n=0,1,2,\ldots,N-1$ (see Fig. 1).

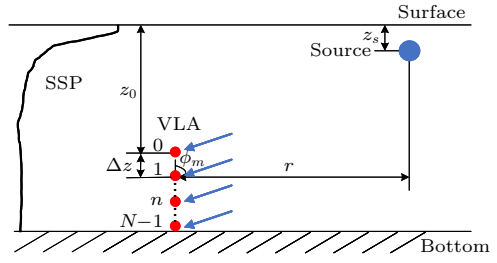


Fig. 1. Schematic of vertical array reception. SSP = sound speed profile.

According to normal mode theory, when the target source is located at depth z_s and the range from the vertical array is r, the acoustic pressure signal received by the n-th array element can be expressed as[24]

$$P^n(r) = \frac{\mathrm{j}\exp(-\mathrm{j}\pi/4)}{\sqrt{8\pi}\rho(z_s)} \times \sum_{m=1}^{M} u_m(z_s)u_m(z_0+n\Delta z)\frac{\exp(\mathrm{j}k_{rm}r-\alpha_m r)}{\sqrt{k_{rm}r}}, \quad (1)$$

where M denotes the total number of normal modes. $u_m(\cdot)$, k_{rm}, and α_m represent the eigenfunction, horizontal wavenumber, and attenuation coefficient of the m-th mode, respectively. ρ denotes the water density.

By applying beamforming to the received sound field of the vertical array, a function of the scanning angle ϕ as the independent variable is obtained, which characterizes the sound field elevation structure:

$$B_p(\cos\phi) = \frac{1}{N}\sum_{n=0}^{N-1} p^n(r)\exp(\mathrm{j}k(z_0+n\Delta z)\cos\phi). \quad (2)$$

Here, the horizontal wavenumber is given by $k=2\pi f/c$, where c is the sound speed.

By substituting Eq. (1) into Eq. (2), one obtains

$$B_p(\cos\phi) = \sum_{m=1}^{M}\left\{E\cdot u_m(z_s)\frac{\exp(\mathrm{j}k_{rm}r-\alpha_m r)}{\sqrt{k_{rm}r}}\cdots \times\left[\frac{1}{N}\sum_{n=0}^{N-1}u_m(z_0+n\Delta z)\exp(\mathrm{j}k(z_0+n\Delta z)\cos\phi)\right]\right\}. \quad (3)$$

Here, $E=\mathrm{j}\exp(-\mathrm{j}\pi/4)/\left(\sqrt{8\pi}\rho(z_s)\right)$.

In practical applications, large-aperture vertical arrays often face challenges such as deployment difficulties and array tilting. To address these, this study adopts a small-aperture vertical array moored to the seabed, with array elements mainly distributed in a narrow region near the bottom. Although the eigenfunction is a global solution determined by the complete depth-dependent sound speed profile, the sound speed variation within the depth range covered by the vertical array is typically small in this study. Therefore, a simplified expression based on a constant sound speed is used to approximate the local eigenfunction. The eigenfunction $u_m(z_0+n\Delta z)$ can be expressed as[23]

$$u_m(z_0+n\Delta z) \simeq A_m^+\exp(\mathrm{j}k_{zm}(z_0+n\Delta z)) + A_m^-\exp(-\mathrm{j}k_{zm}(z_0+n\Delta z)), \quad (4)$$

where A_m^+ and A_m^- denote the amplitudes of the upward and downward plane waves of the m-th mode. k_{zm} is the vertical wavenumber of the m-th mode and satisfies $k_{zm}=k\cos\phi_m$, where ϕ_m is the m-th mode's elevation angle. ϕ_m represents the angle between the modal wavevector and the vertical direction. The Wentzel–Kramers–Brillouin (WKB) approximation[29] links the normal mode theory to ray acoustics, thereby giving ϕ_m a physical meaning analogous to the ray arrival angle.

By substituting Eq. (4) into Eq. (3) and rearranging, Eqs. (5)–(7) are obtained. This shows that the sound field elevation structure is represented as the weighted sum of the mode elevation structures $B_{um}(\cos\phi)$ of all modes, where the modal weighting coefficients are denoted by C_m:

$$B_p(\cos\phi) = \sum_{m=1}^{M} C_m B_{um}(\cos\phi), \quad (5)$$

$$B_{um}(\cos\phi) = \frac{1}{N}\sum_{n=0}^{N-1}\left[u_m(z_0+n\Delta z)\exp\left(\mathrm{j}k(z_0+n\Delta z)\cos\phi\right)\right]$$

$$= \frac{(N-1)\Delta z}{N}\left[A_m^+ \mathrm{e}^{\mathrm{j}k(\cos\phi+\cos\phi_m)\left(z_0+\frac{(N-1)\Delta z}{2}\right)}\right.$$
$$\times \operatorname{sinc}\left(\frac{k(\cos\phi+\cos\phi_m)(N-1)\Delta z}{2\pi}\right)\cdots$$
$$+A_m^- \mathrm{e}^{\mathrm{j}k(\cos\phi-\cos\phi_m)\left(z_0+\frac{(N-1)\Delta z}{2}\right)}$$
$$\left.\times \operatorname{sinc}\left(\frac{k(\cos\phi-\cos\phi_m)(N-1)\Delta z}{2\pi}\right)\right], \quad (6)$$

$$C_m = E\cdot u_m(z_s)\frac{\exp(\mathrm{j}k_{rm}r-\alpha_m r)}{\sqrt{k_{rm}r}}. \quad (7)$$

According to Eqs. (5)–(7), the sound field elevation structure is determined by the mode elevation structure and the modal weighting coefficients. In particular, the mode elevation structure is governed by the array configuration and the underwater acoustic channel, while the modal weighting coefficients depend on the target depth. In a shallow-water negative thermocline waveguide, modal trapping occurs: low-order modes, referred to as trapped modes, are difficult for surface targets to excite, whereas high-order modes, known as free modes, can be excited by both surface and underwater targets.[10] As a result, the weighting coefficients of trapped modes for surface targets are generally smaller than those of underwater targets. The target depth affects the sound field elevation structure through its influence on the modal weighting coefficients, leading to significant differences between the elevation structures of surface and underwater targets.

The following presents a simulation-based analysis of the sound field elevation structure samples for targets located at different depths. The simulations are conducted using the Kraken Acoustic Toolbox.[21] The trial sound speed profile is shown in Fig. 2. The seabed consists of a 23.5 m thick sediment layer, an 800 m thick mudstone layer, and a half-space. The geoacoustic properties of each layer are listed in Table 1.

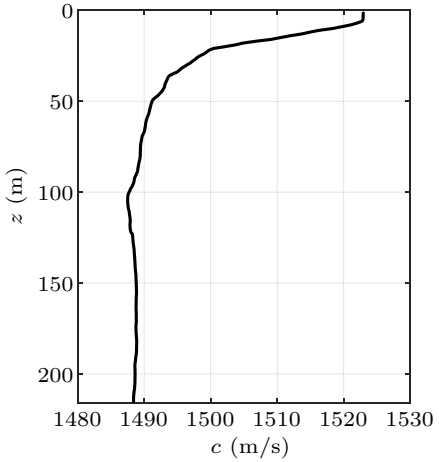


Fig. 2. Trial sound speed profile.

Table 1. Geoacoustic parameters of seabed sediment layers.

Thickness (m)	Sound speed (m/s)	Density (g/cm^3)	Attenuation (dB/kmHz)
23.5	1572.3–1593.0	1.76	0.20
800	1881.0–3245.0	2.06	0.06
∞	5200.0	2.66	0.02

The vertical array fixed to the seabed consists of six elements spaced at 5 m intervals. The array covers depths from 187 m to 212 m, which accounts for approximately 12% of the total water depth. The array scanning angle ϕ is restricted to values in the interval $(0,\pi)$. The target range is 5 km. Based on the relationship between modal phase velocity and the maximum sound speed in the water column, the propagation modes are classified into trapped modes and free modes.[10] Under the waveguide environment considered in this section, modes one to six are identified as trapped modes, and the seventh mode is the lowest-order free mode, referred to as the critical mode. The cosine of the modal elevation angle for the critical mode, denoted by $\cos\phi_m$, is 0.1899. Accordingly, the angular region $(-1,-0.1899]\cup[0.1899,1)$ corresponds to free modes, while the region $(-0.1899,0.1899)$ corresponds to trapped modes.

The sound field elevation structure sample can be represented as a vector

$$\boldsymbol{x} = [x_1, x_2, \ldots, x_k, \ldots, x_K], \quad k = 1,2,\ldots,K, \quad (8)$$

where $x_k = B_p(\cos\phi_k)$ and ϕ_k is the k-th array scanning angle. K denotes the total number of array scanning angles.

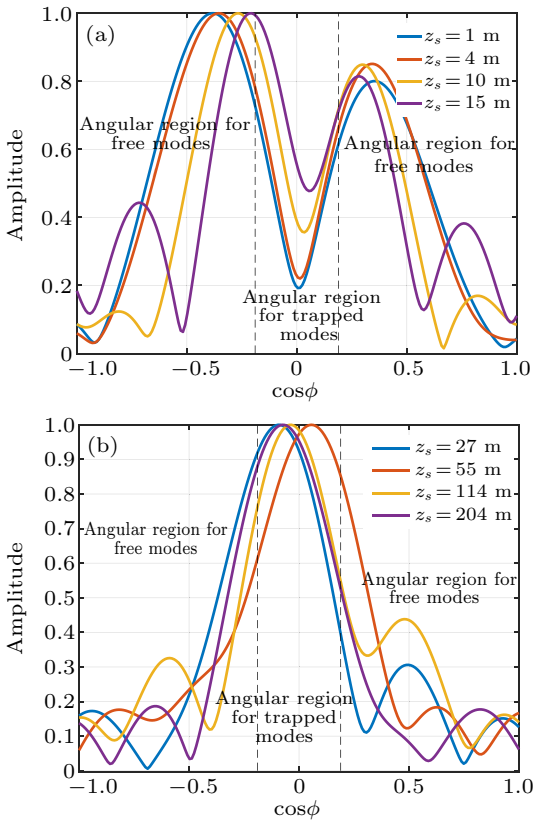


Fig. 3. Normalized sound field elevation structure samples for targets at different depths: (a) surface targets, (b) underwater targets. Note: $\cos\phi$ represents the cosine of the scanning angle.

Min–max normalization is applied to rescale all elements of the sample $\boldsymbol{x}$ into the interval [0,1], yielding the normalized sound field elevation structure sample $\tilde{\boldsymbol{x}}$, given by

$$\tilde{\boldsymbol{x}} = \frac{\boldsymbol{x}-\min(\boldsymbol{x})}{\max(\boldsymbol{x})-\min(\boldsymbol{x})}. \quad (9)$$

Several normalized samples of sound field elevation structures for targets at different depths are shown in Fig. 3.

It can be observed that when the target depth is less than or equal to 15 m, the energy of the sound field elevation structure is mainly distributed in the free-mode angular region. When the target is located at a greater depth, the energy is primarily concentrated in the trapped-mode region. This indicates that the distributions of sound field elevation structures for surface and underwater targets differ in the cosine of the elevation angle, $\cos\phi$. Therefore, the sound field elevation structure is adopted as the input feature for the classification network model.

3. UDA-ResNet-based surface/underwater target classification model

3.1. ResNet classification network

ResNet is widely used to learn the deep features of sound field data[7,16] as it effectively mitigates the vanishing gradient problem in deep network training.[17] In this paper, the ResNet classification network serves as the backbone of the proposed UDA-ResNet model. This section first presents the design of the ResNet classification network.

Considering the characteristics of sound field elevation structure data, a lightweight ResNet10 is adopted as the backbone model for surface/underwater target classification. ResNet10 mainly consists of four residual blocks, each composed of two convolutional layers (Conv) with skip connections to enable direct information flow.

Based on the traditional ResNet framework, 1D convolutional kernels are used to extract features from the input vector $\bar{x} \in R^{1\times K}$, accommodating the vector form of the sound field elevation structure data. Each convolutional layer is followed by batch normalization (BN) and a rectified linear unit (ReLU), ensuring network stability and enhancing nonlinear representational capacity. The process is defined as

$$\Phi(\bar{\boldsymbol{x}}) = \mathrm{ReLU}(\mathrm{BN}(W_2 * \mathrm{ReLU}(\mathrm{BN}(W_1 * \bar{\boldsymbol{x}})))), \tag{10}$$

where W_1 and W_2 are one-dimensional convolutional kernels and $*$ denotes the convolution operation. In the second convolutional layer of each residual block, dropout is applied before the activation layer. This randomly masks part of the features, reducing overfitting to specific patterns.

The residual block performs element-wise addition between the output and the input features, yielding the final output

$$\boldsymbol{z} = \bar{\boldsymbol{x}} + \Phi(\bar{\boldsymbol{x}}). \tag{11}$$

This design helps mitigate the vanishing or exploding gradient problem that may occur in deep networks,[17] while also reducing the computational complexity of the model.

For the surface/underwater target classification problem, the network extracts key information from the input features layer by layer through residual blocks. The information is then compressed into a feature vector by a global average pooling layer. This feature vector contains the target depth-related features learned by the network. Subsequently, it is mapped to the class space through a fully connected (FC) layer and converted into predicted probabilities by the softmax layer. Finally, a decision criterion is defined using hypothesis testing:

$$H_0 : P_0 \geq P_1, \quad H_1 : P_0 < P_1, \tag{12}$$

where P_0 and P_1 represent the probabilities of the sample being predicted as class zero (a surface target) and class one (an underwater target), respectively. If the probability of predicting the sample as a surface target is lower than that for an underwater target, the sample is classified as an underwater target; otherwise, it is classified as a surface target.

The structure of the ResNet classification network is shown in Fig. 4. The input sound field elevation structure sample has a size of $1 \times 1 \times K$, with $K = 200$ set in this case. The relevant parameters of the ResNet classification network are listed in Table 2.

Table 2. Relevant parameters of the ResNet classification network.

Layer name	Number of convolutional layers	Number of kernels, kernel size, stride	Output size
Input layer	1	(64, 1×7, 2)	64×1×50
Residual block #1	1	(64, 1×3, 1)	64×1×50
	1	(64, 1×3, 1)	64×1×50
Residual block #2	1	(128, 1×3, 2)	128×1×25
	1	(128, 1×3, 1)	128×1×25
Residual block #3	1	(256, 1×3, 2)	256×1×13
	1	(256, 1×3, 1)	256×1×13
Residual block #4	1	(512, 1×3, 2)	512×1×7
	1	(512, 1×3, 1)	512×1×7
Average pooling	–	–	512
FC	–	–	2

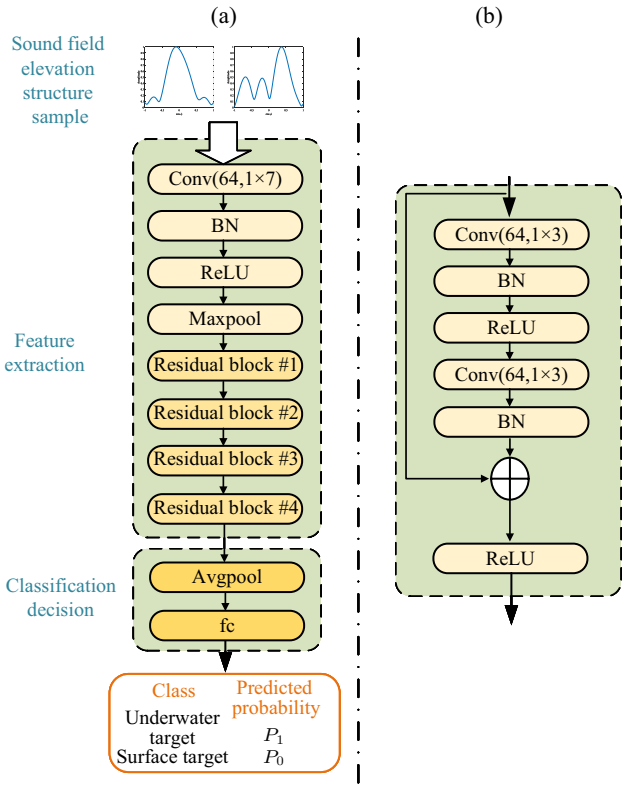


Fig. 4. Structure of the ResNet classification network. (a) ResNet classification network. (b) Residual block.

3.2. UDA-ResNet improved model

The scale and quality of the training data are key factors affecting the performance of the ResNet algorithm. However, due to the complexity and variability of marine environments, the collection of acoustic data is not only challenging but also costly, making it difficult to obtain large numbers of labeled sound field elevation structure samples. ResNet is limited by insufficient training samples, which may lead to overfitting, thereby affecting classification performance in practical applications.

To address these issues, an improved model based on UDA-ResNet is proposed in this paper. The model augments the training set with simulated data, employing adversarial training to extract domain-invariant features common to both simulated and trial data. This ensures that the features extracted by the feature extractor from both the source domain (simulated data) and the target domain (trial data) are indistinguishable to the domain discriminator. As a result, the domain adaptation problem between simulated and trial data is mitigated, leading to improved generalization ability in trial data classification tasks.

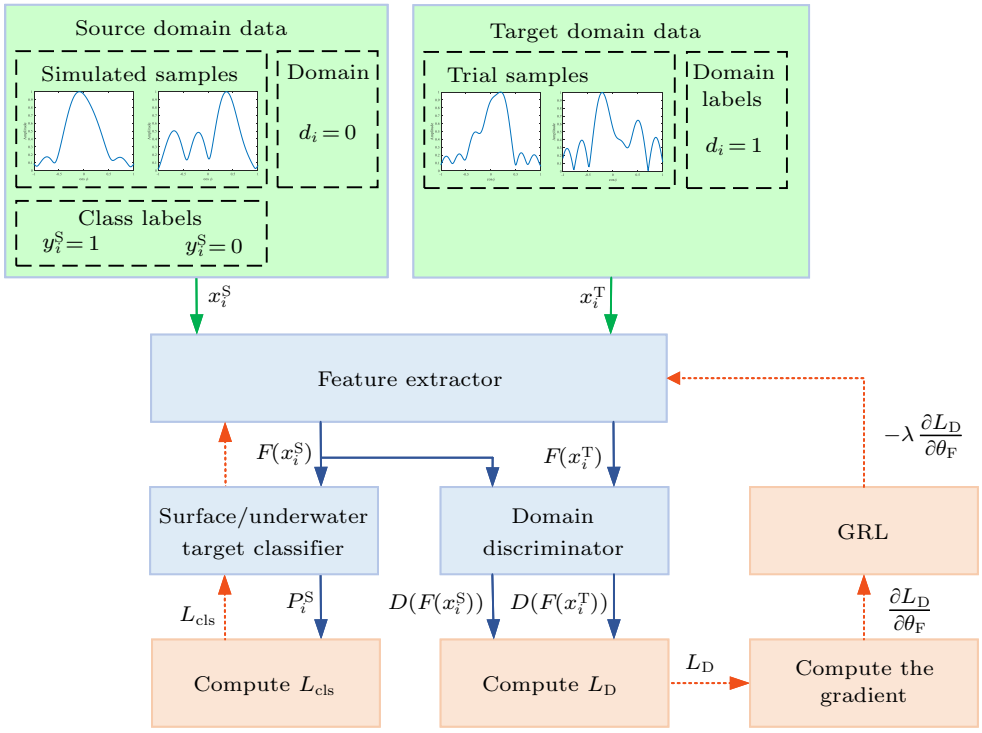


Fig. 5. Block diagram of the UDA-ResNet-based surface/underwater target classification model.

As shown in Fig. 5, the model architecture consists of three parts: the feature extractor, the surface/underwater target classifier, and the domain discriminator.

First, the source domain data is defined as the labeled simulated data $\{x_i^S; y_i^S\}$, where y_i^S is the class label of the i-th sample x_i^S. Here, $y_i^S \in \{0,1\}$. The target domain data is the unlabeled trial data $\{x_i^T\}$. The domain label d_i is set to zero for the source domain data and to one for the target domain data. The feature extractor maps the source domain data and target domain data into the feature representations $F(x_i^S)$ and $F(x_i^T)$.

The surface/underwater target classifier then performs classification based on the source domain features extracted by the feature extractor. The optimization objective of the classifier is to minimize the cross-entropy loss function, thereby improving predictive ability. The cross-entropy loss function is as follows:

$$L_{\text{cls}} = -\frac{1}{N_{\text{S}}}\sum_{i=1}^{N_{\text{S}}}(y_i^{\text{S}}\log(P_i^{\text{S}}) + (1-y_i^{\text{S}})\log(1-P_i^{\text{S}})), \quad (13)$$

where P_i^{S} is the probability that the sample is predicted to belong to class one (underwater target). N_{S} is the number of source domain samples. The loss function in Eq. (13) guides the model to correctly classify the source domain data by comparing the predicted probabilities with the true labels. The relevant parameters of the feature extractor and the surface/underwater target classifier are shown in Table 2.

The domain discriminator receives the output features from the feature extractor and determines whether they originate from the source domain or the target domain. During training, adversarial optimization occurs between the domain discriminator and the feature extractor: the domain discriminator aims to accurately distinguish between samples from the source and target domains, while the feature extractor adjusts its parameters via backpropagation to ensure that its extracted features are indistinguishable to the domain discriminator.

In this context, 'domain-invariant features' refer to the condition where the sound field elevation structure samples from the source and target domains exhibit similar statistical distributions in the feature space, i.e.,

$$\text{div}(P(F(x^{\text{S}})), P(F(x^{\text{T}}))) < \varepsilon. \quad (14)$$

where $P\left(F(x^{\text{S}})\right)$ and $P\left(F(x^{\text{T}})\right)$ represent the probability distributions of the source and target domain features, respectively. Here, $\text{div}(\cdot,\cdot)$ denotes a distribution distance metric, and this study adopts the commonly used Jensen–Shannon divergence (JSD).[30] ε represents the acceptable upper bound of distribution discrepancy. In this context, if the extracted features from the source and target domains exhibit a small JSD (e.g., < 0.03), and the classification accuracy in the target domain reaches a high level (e.g., > 0.8), then these features can be considered domain-invariant. It should be noted that to extract domain-invariant features, adversarial training must meet the following conditions: 1) the target and source domain data must have intrinsic consistency; 2) the source domain must contain sufficient and diverse labeled samples, and the feature extractor must have a strong representation capability; 3) adversarial training should reach equilibrium; otherwise, adjustments such as tuning the adversarial loss weight or modifying the discriminator may be needed.

To achieve this, a gradient reversal layer (GRL)[20] is used to reverse the gradient of the adversarial loss function with respect to the feature extractor parameters θ_{F}. During the optimization process, the adversarial loss is maximized, enabling adversarial training between the feature extractor and the domain discriminator. The adversarial loss function is

$$L_{\text{D}} = -\frac{1}{N_{\text{S}}+N_{\text{T}}}\sum_{i=1}^{N_{\text{S}}+N_{\text{T}}}[(1-d_i)\log(1-(D(F(x_i)))) + d_i\log(D(F(x_i)))], \quad (15)$$

where N_{T} represents the number of target domain samples. $D(F(x_i))$ is the probability that sample x_i is predicted to belong to the target domain by the domain discriminator.

The GRL does not, in itself, affect the forward propagation process, but during backward propagation, it reverses the gradient with a negative scaling factor of $-\lambda$, thereby enabling adversarial training. The definition of the adversarial factor λ is as follows:

$$\lambda = \frac{2}{1+\exp(-\gamma\cdot\text{epoch}/\text{epochs})} - 1, \quad (16)$$

where epoch/epochs represent the current/total training progress. γ controls the growth rate of the adversarial factor, thereby dynamically adjusting the adversarial strength of the GRL. In the early stages, the focus is on the classification task, while in the later stages, the adversarial strength is enhanced to learn domain-invariant features, thus improving the model's generalization ability on the target domain.

The total loss function L_{sum} is defined as the weighted sum of the classification loss and the adversarial loss, as shown below:

$$L_{\text{sum}} = L_{\text{cls}} + a_{\text{D}}L_{\text{D}}, \quad (17)$$

where a_{D} is the adversarial loss weight used to control the strength of adversarial training. The objective of the total loss function is to minimize classification error while using adversarial training to reduce the discrepancy between the feature distributions of the source and target domains. This helps the feature extractor learn domain-invariant features and improves classification performance on the target domain.

4. Performance evaluation

This section first introduces the constructed simulated dataset and the trial dataset. Then, the ResNet classification network, along with the UDA-ResNet improved model, is trained and evaluated for performance in surface/underwater target classification tasks. For comparison, we introduce a classification method based on the trapped–free modal sound field elevation structure energy ratio (the TFMR method),[5] which was recently proposed by our research team. This method uses a binary classifier to distinguish between surface and underwater targets. It is used as a benchmark for comparing the classification performances of the models proposed in this study.

4.1. Simulated dataset

Data collection is conducted through simulation experiments based on a vertical array, with sound field elevation structure samples used as the model input. The environmental parameters used in the simulation are as follows: seven sound speed profiles (SSP#0–SSP#6) are set, as shown in Fig. 6; two seabed types are configured: seabed #1 is a three-layer seabed, and seabed #2 is a half-space seabed. The relevant parameters are listed in Table 3.

The targets to be classified include both surface and underwater targets. The depths of the surface targets are randomly selected between 1 m and 15 m, while the depths of the underwater targets range from 16 m to 216 m.

Different signal-to-noise ratios (SNRs) and target range parameters are used to construct each dataset, with the aim of evaluating whether the model can learn general features rather than just memorizing the training samples. The detailed parameters of each dataset are shown in Table 4. Other simulation conditions are consistent with those described in Section 2.

Table 4 presents two training sets. The difference between them is as follows: the environmental parameters of training set #1 match those of the validation and test sets, while the parameters of training set #2 differ. Training set #2 is used to evaluate classification performance under conditions of domain mismatch.

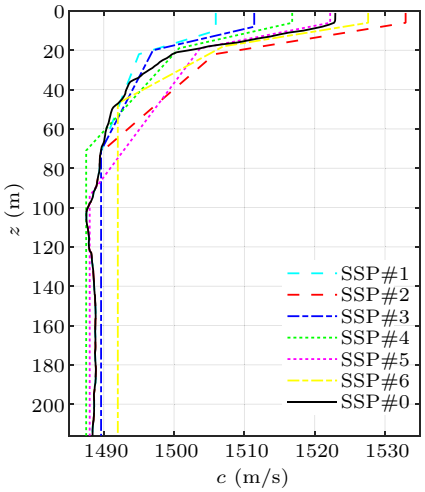


Fig. 6. Sound speed profiles.

Table 3. Relevant parameters of the seabed.

Type	Thickness (m)	Sound speed (m/s)	Density (g/cm^3)	Attenuation (dB/kmHz)
	23.5	1572.3–1593.0	1.76	0.20
Seabed #1	800	1881.0–3245.0	2.06	0.06
	∞	5200.0	2.66	0.02
Seabed #2	∞	1700	1.80	0.20

Table 4. Relevant parameters used to generate the training, validation, and test sets.

Parameter	Unit	Training set #1	Training set #2	Validation set	Test set
SNR	dB	10:2:20	20	11:2:19	0:5:35
Range	km	3.0:0.012:9.0		3.0:0.1:9.0	2.5:1.0:9.5
Surface target depth	m	1–15, randomly selected 40 depths			
Underwater target depth	m	16–216, randomly selected 40 depths			
Sound speed profile	–	SSP#0	SSP#1–SSP#6	SSP#0	SSP#0
Seabed	–	seabed #1	seabed #2	seabed #1	seabed #1

Note: The notation "10:2:20" in the table is used to describe the value span and step size of a parameter. Here, it indicates that the SNR spans from 10 dB to 20 dB, with a step size of 2 dB.

4.2. Trial dataset

The trial data were collected from the SWellEx-96 experiment (Fig. 7), conducted off the coast of San Diego, California.[22] The measured sound speed profile corresponds to SSP#0 in Fig. 6. In the trial, the source ship traveled northward along an isobath at a speed of 2.5 m/s, towing two acoustic sources at depths of 9 m and 54 m. These sources transmitted continuous wave signals at multiple discrete frequencies. The vertical array comprised 21 elements, spanning depths from 94.125 m to 212.25 m, with an inter-element spacing of approximately 5.6 m.

Array elements 16 to 21 are selected to form a six-element array with an aperture of 28.13 m. The received data are segmented into 10 s frames with a 75% overlap. Surface and underwater target data are extracted separately to compute their corresponding sound field elevation structures. A total of 957 samples for surface targets and 802 samples for underwater targets are obtained, as shown in Fig. 8.

4.3. Performance metrics

The classification accuracies for surface and underwater targets, denoted by $\mathrm{Acc_s}$ and $\mathrm{Acc_d}$, respectively, are used as performance evaluation metrics for the model:

$$\mathrm{Acc_s} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \tag{18}$$

$$\mathrm{Acc_s} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{19}$$

where TN is the number of surface target samples correctly predicted to be surface targets, FP is the number of surface target samples incorrectly predicted to be underwater targets, and TP and FN denote the numbers of underwater target samples correctly and incorrectly classified, respectively.

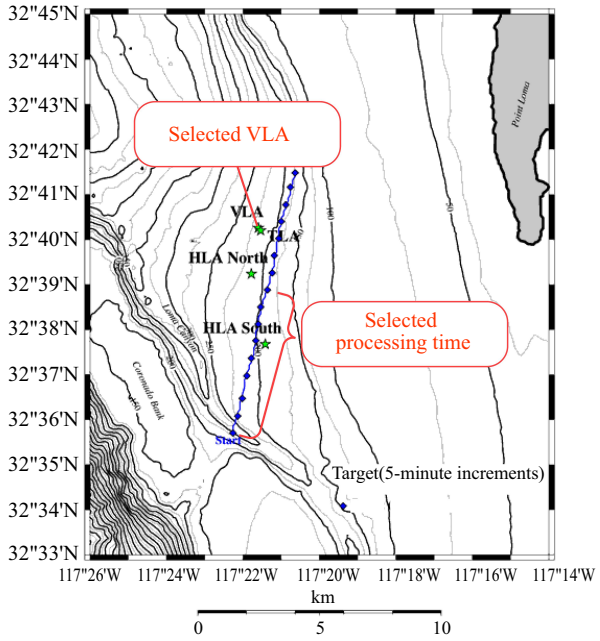


Fig. 7. SWellEx-96 experimental scenario.

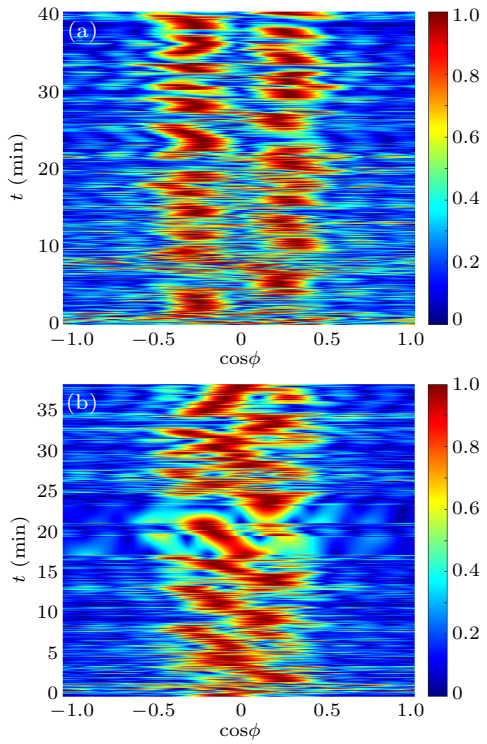


Fig. 8. Normalized sound field elevation structure samples from trial data. (a) Surface target. (b) Underwater target. Note: $\cos\phi$ represents the cosine of the scanning angle.

4.4. Performance analysis of the ResNet classification network

Since the ResNet classification network serves as the backbone of the UDA-ResNet model, its classification performance is first trained and evaluated using the simulated dataset.

The network weights are initialized using the Kaiming initialization method. Training is performed using the Adam optimizer with an initial learning rate of 10^{-4}. A learning rate scheduler is applied to facilitate efficient learning, and early stopping is used to prevent overfitting.

The computational platform is equipped with an NVIDIA RTX 4090D GPU (24 GB VRAM), a 16-core Intel Xeon Platinum 8481C CPU, and 80 GB of RAM. All experiments are conducted using the PyTorch 2.5.1 deep learning framework running in a Python 3.12 environment, with GPU acceleration enabled via CUDA 12.4.

The ResNet classification network is trained on training set #1, while the validation set is used for model tuning. The parameters from the epoch with the lowest validation loss are retained. The training and validation loss curves are shown in Fig. 9.

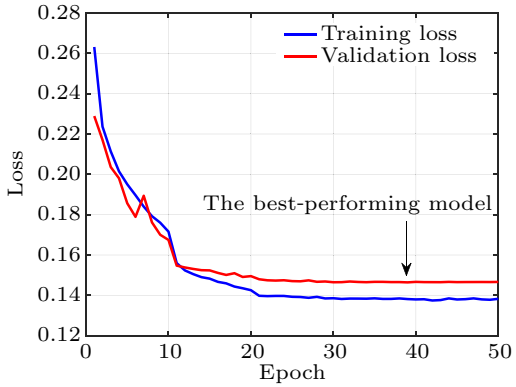


Fig. 9. Training and validation loss curves.

As illustrated, both training and validation losses gradually converge with increasing epochs, and the validation loss closely follows the training loss. With the current training sample size, no noticeable overfitting is observed during the training of the ResNet.

Figure 10(a) shows the t-distributed stochastic neighbor embedding (t-SNE) distributions of 500 randomly selected samples of underwater and surface targets from the test set. Figure 10(b) presents the corresponding deep features extracted by residual block #4 of the ResNet classification network. In each plot, a smaller overlap between the two distributions indicates less difficulty in distinguishing surface targets from underwater targets.

In the t-SNE visualization of the test samples, surface and underwater targets exhibit substantial overlap, indicating blurred class boundaries. Notably, the t-SNE visualization of deep features extracted by ResNet shows better separation between surface and underwater targets in the feature space. This indicates that the ResNet classification network effectively ex-

tracts deep features and provides an advantage in classification performance.

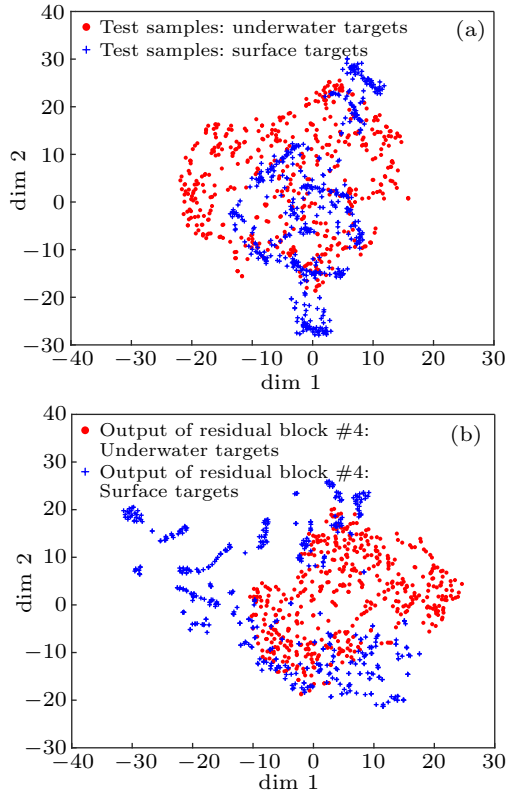


Fig. 10. The t-SNE distributions. (a) Test samples. (b) Deep features of corresponding samples extracted by residual block #4.

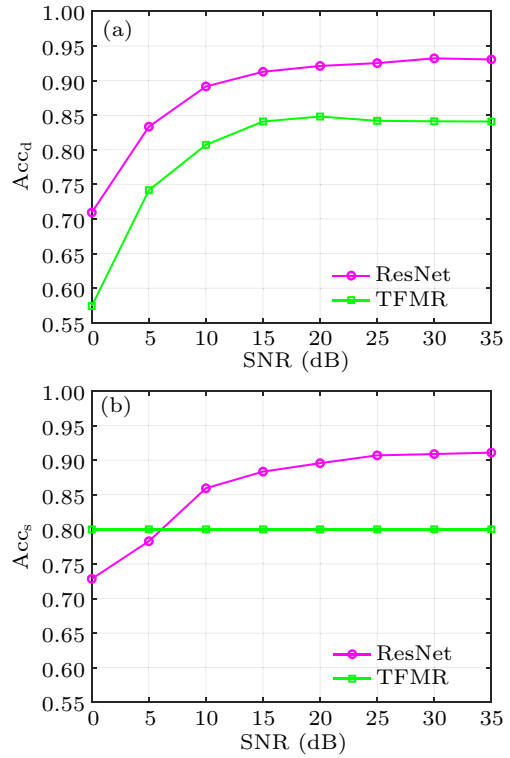


Fig. 11. Classification accuracy at different SNRs. (a) Underwater targets. (b) Surface targets.

Figure 11 presents the classification accuracy of ResNet on the test set and a comparison with the TFMR method. In this evaluation, the TFMR method adopts a constant false alarm rate detection strategy, where the classification accuracy for surface targets is fixed at 0.80. Based on this setting, the reported metric reflects the classification accuracy for underwater targets.

The ResNet classification network performs well in both underwater and surface target classification. At SNRs above 5 dB, the classification accuracy is significantly higher than that of the TFMR method. At an SNR of 20 dB, classification accuracies reach 0.92 for underwater targets and 0.89 for surface targets, compared to 0.85 and 0.80 with the TFMR method. Overall, the ResNet classification network exhibits strong classification capability under high SNR conditions and maintains robustness under low SNRs.

The classification performance of the ResNet classification network is further evaluated under conditions of environmental mismatch. Ocean acoustic channels are complex and highly variable. As a result, it is often difficult to obtain accurate and complete environmental parameters, which frequently leads to mismatch issues in classification. To assess generalization ability across different environments, the ResNet classification network is trained using training set #2, in which the sound speed profiles and seabed parameters differ from those in the validation and test sets. A comparative analysis is also conducted with the TFMR method. The detailed environmental parameters are listed in Table 5.

Table 5. Environmental conditions used by ResNet and TFMR at each stage.

Method	Stage	Sound speed profile	Seabed
ResNet	training set #2	SSP1#–SSP#6	seabed #2
	validation set	SSP#0	seabed #1
	test set	SSP#0	seabed #1
TFMR	classification features	SSP#1	seabed #2
	construction	SSP#2	seabed #2
	data to be classified	SSP#0	seabed #1

The classification accuracies of the ResNet and TFMR methods under environmental mismatch conditions are presented in Fig. 12.

As shown, the ResNet classification network maintains robust performance under mismatch conditions. Specifically, classification accuracies for underwater and surface targets reach 0.87 and 0.83, respectively, outperforming the TFMR method. This is mainly because the ResNet classification network can learn discriminative features from diverse training data, thereby maintaining robustness against variations in environmental parameters.

Next, the performance of the ResNet classification network is evaluated under different training sample sizes. Training samples are randomly selected from training set #2. The classification accuracy is shown in Fig. 13. In this figure, the

dotted lines represent the classification accuracies of ResNet trained using the complete training set #2 (120240 samples). The dashed lines represent the classification accuracy of the TFMR method, with results obtained from Fig. 12.

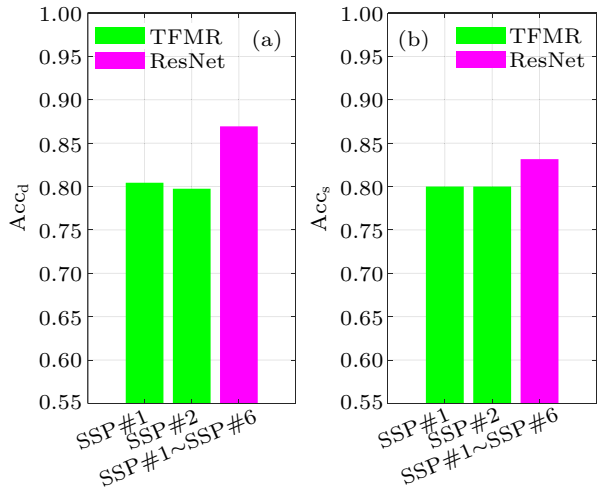


Fig. 12. Classification accuracy under environmental mismatch conditions. (a) Underwater targets. (b) Surface targets.

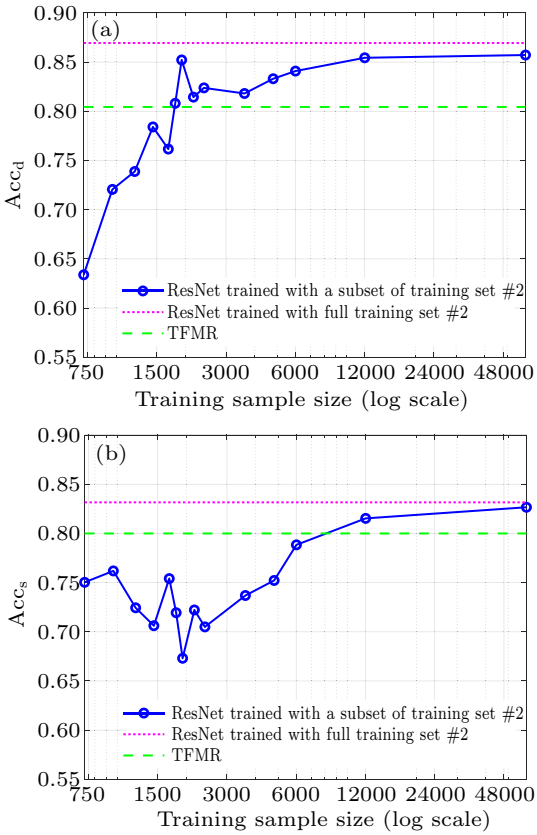


Fig. 13. Classification accuracy versus training sample size. (a) Underwater targets. (b) Surface targets.

Overall, there is a decline in the classification accuracy of ResNet as the training sample size decreases. In certain cases, it even falls below that of the TFMR method. This is because limited training data cannot effectively support the training of ResNet. As a result, ResNet finds it difficult to fully learn discriminative features, which restricts generalization ability. By comparison, the TFMR method relies on a binary classifier and does not require large-scale training. In summary, the performance of the ResNet classification network is significantly influenced by the training sample size. It shows a performance advantage with large training datasets, but this advantage diminishes under limited sample sizes.

4.5. Performance analysis of the UDA-ResNet improved model

The size of the trial data in the SWellEx-96 experiment is limited, with only a few hundred sound field elevation structure samples available. This is not sufficient to directly train the ResNet classification network. Therefore, in this section, training set #2 from the simulated dataset is used as the source domain, and the trial dataset is used as the target domain. The classification performance of the UDA-ResNet model is then evaluated using the limited trial data.

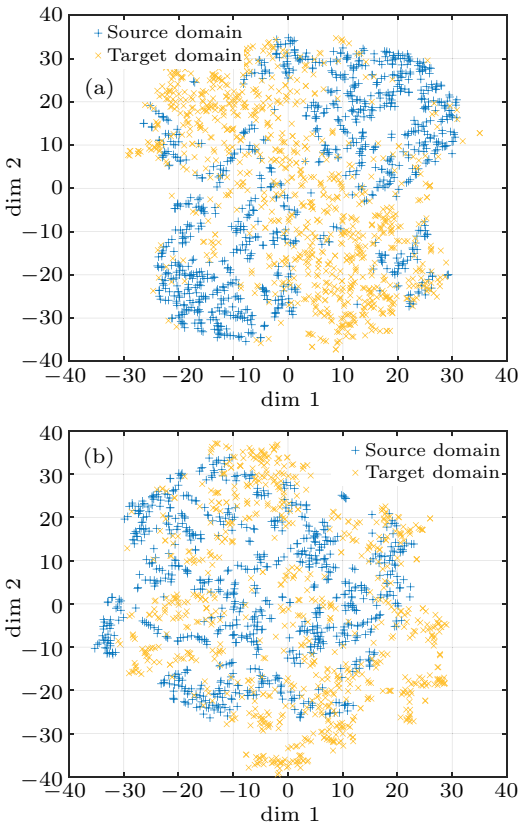


Fig. 14. The t-SNE visualizations of source domain and target domain data. (a) Surface targets. (b) Underwater targets.

First, an analysis is conducted on the feature distribution differences between the source and target domains. Figure 14(a) presents the t-SNE distribution of the surface target samples from the target and source domains, while Fig. 14(b) shows the distribution for underwater target samples. In each plot, a larger overlap area between the two distributions indicates smaller feature discrepancies between the domains. This suggests lower difficulty in distinguishing surface and underwater targets in the target domain.

There are some differences between the t-SNE distributions of the source and target domain data. This suggests that features learned solely from the source domain do not generalize well to classification in the target domain. Therefore, adversarial training with UDA-ResNet is employed to extract domain-invariant features and improve classification performance in the target domain.

To optimize the UDA-ResNet structure, two domain discriminators with different numbers of FC layers are designed, as shown in Table 6. Different subscripts are used to distinguish between them. For the discriminator with more layers, a higher interlayer dropout rate is applied to enhance robustness. The growth rate of the adversarial factor γ is set to ten, and the adversarial loss weight a_{D} is set to 0.3.

Table 6. Parameters of the domain discriminator.

Domain discriminator	Number of FC layers	Dropout
D-fc-3	3	0.4
D-fc-5	5	0.6

Next, MMD-ResNet, CORAL-ResNet, ResNet, and TFMR are employed as baseline methods for comparative analysis. The MMD-ResNet and CORAL-ResNet methods build on ResNet. MMD-ResNet minimizes the mean difference[27] between the source and target domains, while CORAL-ResNet minimizes the covariance difference.[28] These approaches aim to reduce domain discrepancy and enhance classification performance in the target domain.

The environmental conditions used at each stage are listed in Table 7. Since it is often difficult to accurately obtain environmental parameters for trial data in practical applications, the source domain is set to training set #2, which has different environmental parameters from the target domain, to better simulate realistic scenarios. For TFMR, the environmental parameters referenced when constructing the classification features are consistent with those of the trial dataset. This implies that TFMR faces a lower classification difficulty compared with UDA-ResNet.

Table 7. Environmental conditions used by different methods at each stage.

Method	Stage	Sound speed profile	Seabed
UDA-ResNet\MMD-ResNet\CORAL-ResNet	source domain: training set #2	SSP1#–SSP#6	seabed #2
	target domain: trial dataset	SSP#0	seabed #1
ResNet	training set: training set #2	SSP1#–SSP#6	seabed #2
	test set: trial dataset	SSP#0	seabed #1
TFMR	classification features construction	SSP#0	seabed #1
	data to be classified: trial dataset	SSP#0	seabed #1

To visually demonstrate the advantages of employing domain-invariant features, the JSD between the source and target domain data is calculated, both at the input level and after network-based high-level feature extraction. Smaller JSD values indicate higher similarity between the two distributions. The JSD values are shown in Table 8. Furthermore, Figs. 15–18 illustrate the t-SNE visualizations of the extracted high-level features.

Table 8. JSD between source and target domain data at the input and high-level features.

JSD value	Surface targets	Underwater targets
Input data	0.061	0.041
ResNet	0.045	0.034
MMD-ResNet	0.037	0.030
CORAL-ResNet	0.035	0.032
UDA-ResNet	0.025	0.028

Compared with the other baseline methods, UDA-ResNet yields smaller JSD values between the high-level features of the source and target domains, and the t-SNE distributions exhibit higher consistency. From a physical perspective, received signals in sea trials are affected by environmental factors, background noise, and experimental setup, resulting in distributional discrepancies between the target and source sound field elevation structure data. UDA-ResNet employs adversarial training to dynamically adjust network parameters, guiding the feature extractor to learn discriminative features essential for surface/underwater target classification while suppressing domain-specific interference unrelated to target categories, thereby enhancing its adaptation capability in the target domain. Although MMD-ResNet and CORAL-ResNet reduce domain shift by aligning global statistics such as the distribution center or shape between source and target domains, they lack dynamic adaptation mechanisms. As a result, their performance is less effective than UDA-ResNet when handling complex domain discrepancies in this context.

The classification accuracies obtained using the trial data are presented in Table 9.

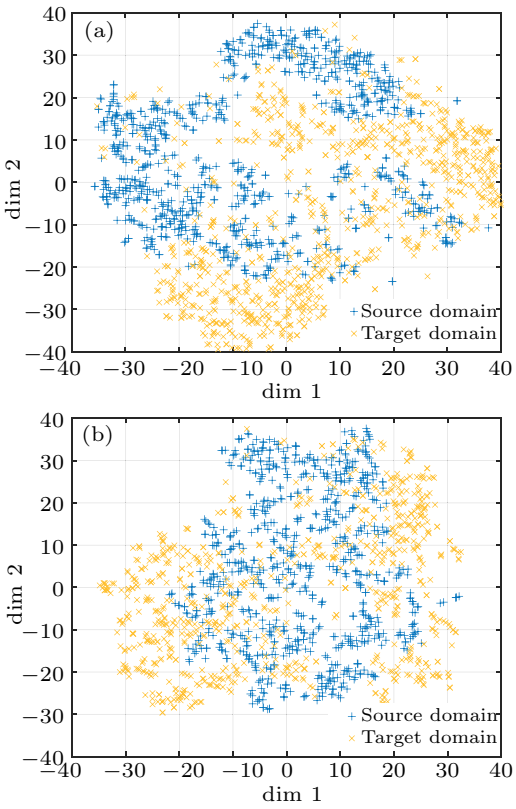


Fig. 15. ResNet: t-SNE visualizations of high-level features extracted from the source and target domains. (a) Surface targets. (b) Underwater targets. MMD-ResNet: t-SNE visualizations of high-level features extracted from the source and target domains. (a) Surface targets. (b) Underwater targets.

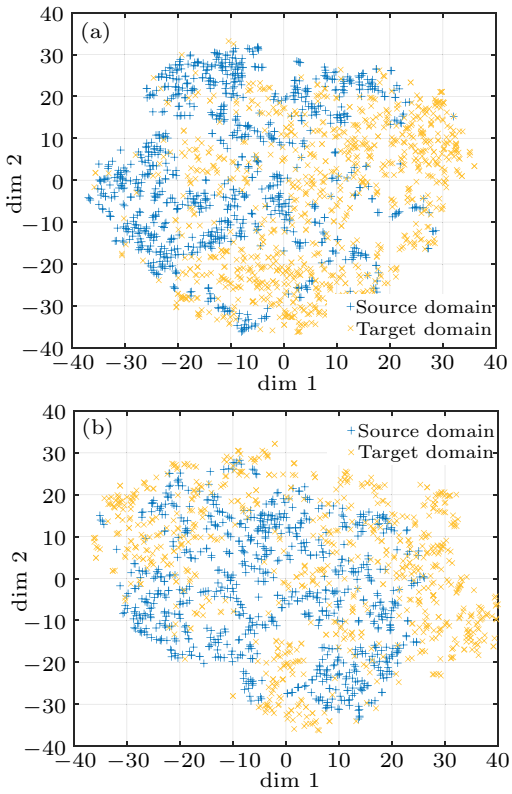


Fig. 16. MMD-ResNet: t-SNE visualizations of high-level features extracted from the source and target domains. (a) Surface targets. (b) Underwater targets.

As shown in the results, UDA-ResNet achieves superior classification performance compared to baseline methods. For surface targets, ResNet achieves an accuracy of

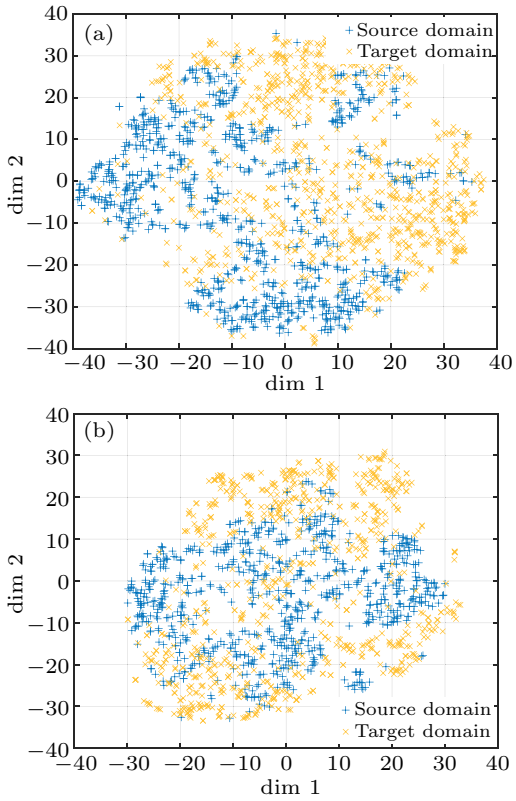


Fig. 17. CORAL-ResNet: t-SNE visualizations of high-level features extracted from the source and target domains. (a) Surface targets. (b) Underwater targets.

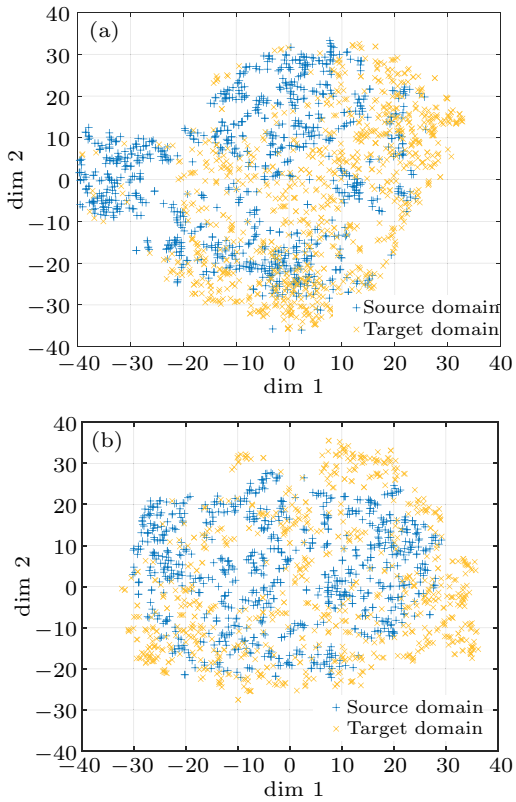


Fig. 18. UDA-ResNet: t-SNE visualizations of high-level features extracted from the source and target domains. (a) Surface targets. (b) Underwater targets.

only 0.56 under limited trial samples, whereas UDA-ResNet (D-fc-5) reaches 0.82. MMD-ResNet and CORAL-ResNet yield only marginal improvements. For underwater targets, ResNet yields an accuracy of 0.78, while UDA-ResNet (D-fc-3) reaches 0.84, with MMD-ResNet and CORAL-ResNet performing comparably. Compared to underwater targets, the trial signals for surface targets are more susceptible to surface boundary effects and environmental interference. This results in a larger discrepancy in the sound field elevation structure between the source and target domains. Therefore, MMD-ResNet and CORAL-ResNet perform less effectively in handling such complex domain discrepancies and fail to significantly improve classification performance in the target domain. In summary, UDA-ResNet mitigates the domain adaptation problem between simulated and trial data. Moreover, it reduces the reliance of the ResNet classification network on large-scale training data and remains effective even under limited trial data conditions.

Table 9. Classification accuracies obtained using the trial data.

Classification accuracy	Surface targets	Underwater targets
TFMR	0.80	0.73
ResNet	0.56	0.78
MMD-ResNet	0.64	0.83
CORAL-ResNet	0.62	0.82
UDA-ResNet (D-fc-3)	0.81	0.84

4.6. Influence of thermocline parameters on UDA-ResNet classification performance

The difference in sound field elevation structures between surface and underwater targets arises from the phenomenon of mode trapping, which typically occurs in shallow-water negative thermocline waveguides. Such waveguides are characterized by parameters including thermocline depth position, thermocline thickness, and so on. To analyze their impact, simulated experiments are conducted to generate target domain data under different thermocline parameters. Sound speed profiles with different thermocline depth positions and thicknesses are configured, as illustrated in Fig. 19. For each profile, 1000 sound field elevation structure samples are generated for both surface and underwater targets. The corresponding test classification accuracies are presented in Table 10.

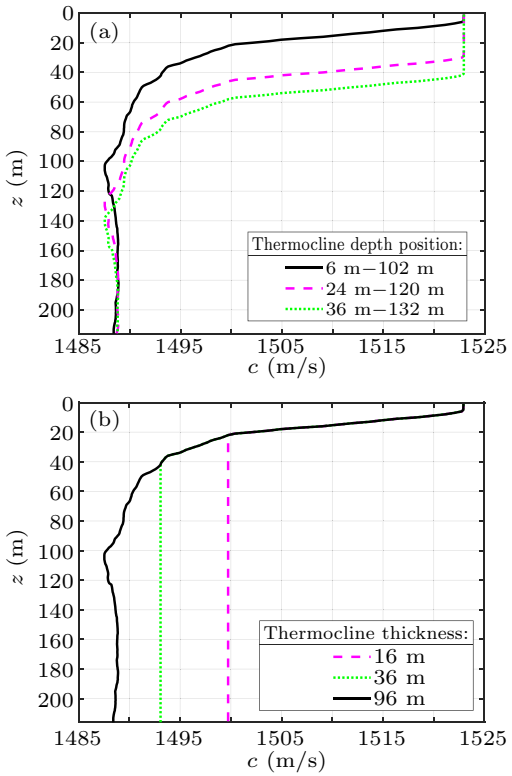


Fig. 19. Sound speed profiles with different thermocline parameters. (a) Different thermocline depth positions. (b) Different thermocline thicknesses.

Table 10. Test classification accuracy for different thermocline parameters.

Thermocline parameter		Surface targets	Underwater targets
Thermocline depth position	6 m–102 m	0.89	0.85
	24 m–120 m	0.92	0.83
	36 m–132 m	0.93	0.79
Thermocline thickness	16 m	0.75	0.69
	36 m	0.87	0.81
	96 m	0.89	0.85

The results show that for UDA-ResNet, increasing the thermocline depth position slightly improves the classification accuracy of surface targets while reducing that of underwater targets. Additionally, increasing the thermocline thickness improves the classification performance for both target types. This is because a deeper thermocline makes it harder for low-order modes to penetrate the high-speed layer near the surface, making surface target features more distinguishable, whereas shallow underwater targets may resemble surface targets. A thicker thermocline increases the sound speed difference across the layer, enhancing the trapping effect and the differences in sound field elevation structure between surface and underwater targets, thus benefiting classification.

5. Conclusions and perspectives

This paper proposes a UDA-ResNet classification model to address the problem of surface/underwater target classification under limited sample sizes using sound field elevation structure data. The model incorporates simulated data to assist

training and extracts domain-invariant features shared between simulated and trial data, thereby reducing the reliance on large amounts of labeled trial data. The classification performance of the proposed method is evaluated through the processing and analysis of simulated and trial datasets. The main conclusions are as follows:

1. Under the hydroacoustic conditions of a shallow-water negative thermocline waveguide, a ResNet-based classification network is constructed using the sound field elevation structure. By progressively stacking residual blocks, the network effectively captures multilevel features from the sound field elevation structure data. Experimental results show that it achieves strong classification capability at high SNRs and exhibits robustness under low SNR conditions. Moreover, when there is an environmental mismatch between the training and testing sets, training with a dataset that incorporates diverse hydrological conditions enables the network to maintain high classification accuracy.

2. To address the generalization issue of the ResNet classification network under limited sample sizes, a UDA-ResNet model is developed based on ResNet. An unsupervised domain adaptation strategy is employed, where simulated sound field elevation structure data are introduced to assist the training process. Domain-invariant features are extracted from both simulated and trial data through adversarial training, which incorporates a domain discriminator and a gradient reversal mechanism. This approach helps mitigate the degradation in classification performance caused by domain distribution discrepancies. Experimental results obtained for trial data show that UDA-ResNet reduces reliance on large-scale labeled trial datasets while remaining effective in distinguishing between surface and underwater targets under limited sample sizes.

It should be noted that this study is based on single-frequency sound field elevation structure data, whereas broadband data contain richer information related to target depth. Therefore, future research will focus on utilizing broadband sound field elevation structure data in combination with deep learning algorithms derived from computer vision. This approach has the potential to further enhance classification performance.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62471024 and 62301183) and the Open Research Fund of Hanjiang Laboratory (KF2024001).

References

[1] Cheng Y S, Qiu J X, Liu Z and Li H T 2019 *Appl. Acoust.* **38** 653
[2] Wang D Z and Shang E C 2013 *Underwater Acoustics* 2nd Edn. (Beijing: Science Press) pp. 124–130, 345–347 (in Chinese)
[3] Yang T 2015 *J. Acoust. Soc. Am.* **138** 1678
[4] Li X B, Sun C and Liu X H 2022 *Acta. Phys. Sin.* **71** 233 (in Chinese)
[5] Miao Y X, Fu J, Zou N, Zhao C P and Dong W F 2025 *Appl. Acoust.* **230** 110412
[6] Zheng G Y and Zhu F W 2021 *Acoust. Aust.* **49** 105
[7] Li Z, Gong L J and Li N S 2024 *Appl. Ocean. Res.* **152** 104201
[8] Cao X, Togneri R, Zhang X and Yu Y 2018 *IEEE. Sens. J.* **14** 1
[9] Emmetiere R, Bonne J, Cristol X, Géhant M and Chonavel T 2019 *IEEE. J-STSP.* **13** 185
[10] Conan E, Bonnel J, Nicolas B and Chonavel T 2017 *J. Acoust. Soc. Am.* **142** 2776
[11] Guo X W, Zheng G Y and Yan Q 2022 *Appl. Acoust.* **198** 108985
[12] Cheng Y S, Li Z Z and Qiu J X 2018 *Underwater Acoustic Target Recognition* (Beijing: Science Press) pp. 12–14 (in Chinese)
[13] Choi J, Choo Y and Lee K 2019 *Sensors* **19** 3492
[14] Luo X W, Qin X M, Wu Z Y, Yang F N, Wang M W and Shang J H 2019 *IEEE Access* **7** 98331
[15] Zhou X Y, Yang K D and Duan R 2019 *IEEE. Signal. Proc. Lett.* **26** 1378
[16] Niu H Q, Gong Z X, Ozanich E, Gerstoft P, Wang H B and Li Z L 2019 *J. Acoust. Soc. Am.* **146** 211
[17] He K M, Zhang X Y, Ren S Q and Sun J 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 26–July 1, 2016, Las Vegas, USA, p. 770
[18] Gao Y J, Chen Y C, Wang F Y and He Y L 2020 *IEEE 5th International Conference on Image, Vision and Computing (ICIVC)*, July 10–12, 2020, Beijing, China, p. 215
[19] Liu F, Ding H, Li D H, Wang T, Luo Z L and Chen L 2021 *OES China Ocean Acoustics (COA)*, July 14–17, 2021, Harbin, China, p. 992
[20] Ganin Y and Lempitsky V 2015 *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, July 6–11, 2015, Lille, France, p. 1180
[21] Porter M B 1991 *The KRAKEN Normal Mode Program* (La Spezia: SACLANT Undersea Research Center) Technical Report SM-245
[22] Information and data for the SWellEx96 experiment
[23] Lee S and Makris N C 2006 *J. Acoust. Soc. Am.* **119** 336
[24] Jensen F B, Kuperman W A, Porter M B and Schmidt H 2011 *Computational Ocean Acoustics* (New York: Springer) pp. 338–341
[25] Li S Y, Cheng L, Li J, Wang Z C and Li J L 2024 *J. Acoust. Soc. Am.* **155** 3410
[26] Cheng L, Ji X Y, Zhao H F, Li J L and Xu W 2022 *J. Acoust. Soc. Am.* **151** 269
[27] Tzeng E, Hoffman J, Zhang N, Saenko K and Darrell T 2014 arXiv 1412.3474
[28] Sun B, Feng J and Saenko K 2016 *Proceedings of the AAAI conference on artificial intelligence*, February 12–17, 2016, Phoenix, Arizona USA, p. 2058
[29] Brekhovskikh L M and Lysanov Y P 1990 *Fundamentals of Ocean Acoustics* (New York: American Institute of Physics Press/ Springer) pp. 131–137
[30] Lin J 2002 *IEEE T. Inform. Theory* **37** 145